

# Voice2Picto : un système de traduction automatique de la parole vers des pictogrammes



CORIA  
TALN  
RJCR | RECITAL  
2023

Cécile Macaire<sup>1</sup>, Emmanuelle Esperança-Rodier<sup>1</sup>, Didier Schwab<sup>1</sup>, Benjamin Lecouteux<sup>1</sup>

<sup>1</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

<sup>1</sup>first.last@univ-grenoble-alpes.fr

## Voice2Picto

- **Système de traduction**, qui, à partir de l'oral, propose une séquence de pictogrammes correspondants.
- Basé sur les technologies du **traitement automatique du langage naturel**.

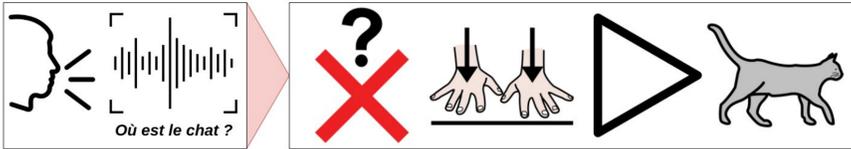


Figure 1 : Schéma simplifié du système de traduction avec comme phrase exemple : « Où est le chat ? ».

## Quels objectifs ?

La **CAA** (Communication Alternative et Augmentée) est mise en place lorsqu'une personne ne peut utiliser les canaux traditionnels de communication (parole, gestes, langue des signes).  
➢ Utilisation de **pictogrammes**, image représentant un concept plus ou moins concret.

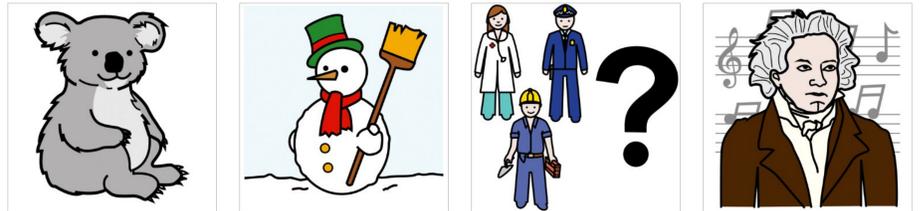


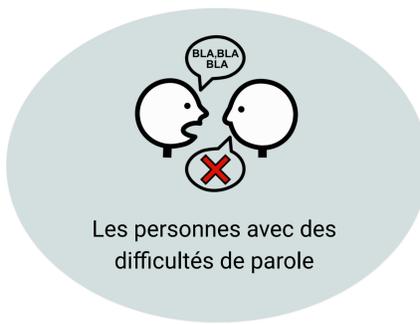
Figure 2 : Quelques exemples de pictogrammes tirés de la banque ARASAAC avec, de gauche à droite, « koala », « bonhomme de neige », « Quel est ton métier ? » et « Beethoven ».

Mais la prise en main des outils de CAA est **longue et difficile** [2] :

- Temps d'adaptation et d'apprentissage nécessaires,
- Manque de soutien institutionnel,
- Lenteurs administratives.

Pourtant, la **CAA a un impact social positif** pour les personnes en situation de handicap [3] : réduction du stress, amélioration de l'autonomie et de l'état de santé, meilleure sérénité.

**Deux objectifs** / Améliorer l'accès à la communication pour... :



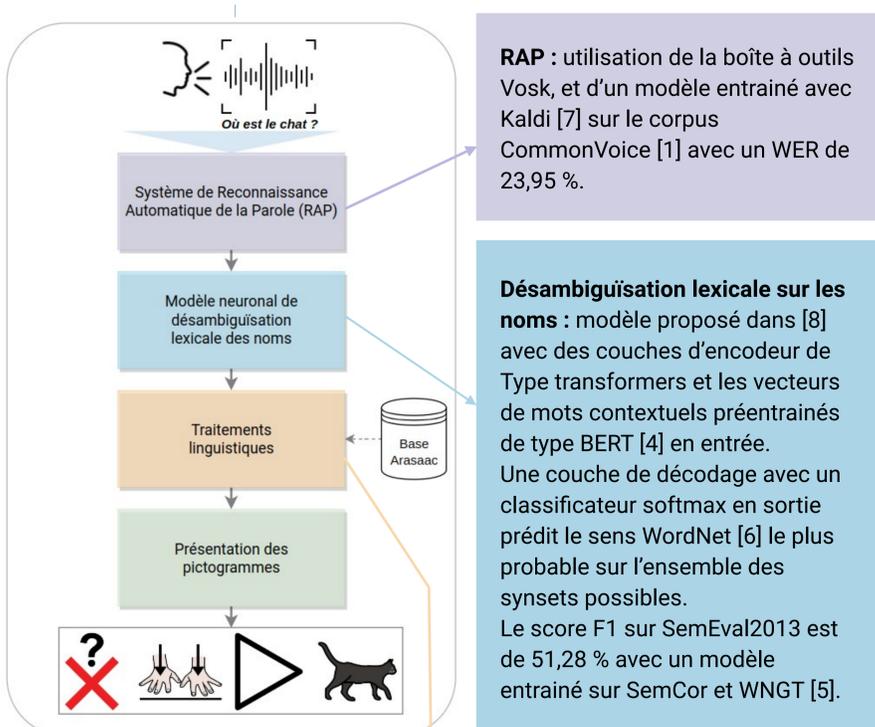
Les personnes avec des difficultés de parole



Les personnes allophones dans un contexte d'urgence médicale

...en proposant un outil aux services hospitaliers et aux familles sans connaissances préalables des pictogrammes de communiquer auprès de ces deux groupes.

## Quelle architecture ?



**RAP** : utilisation de la boîte à outils Vosk, et d'un modèle entraîné avec Kaldi [7] sur le corpus CommonVoice [1] avec un WER de 23,95 %.

**Désambiguïsation lexicale sur les noms** : modèle proposé dans [8] avec des couches d'encodeur de Type transformers et les vecteurs de mots contextuels préentraînés de type BERT [4] en entrée. Une couche de décodage avec un classificateur softmax en sortie prédit le sens WordNet [6] le plus probable sur l'ensemble des synsets possibles. Le score F1 sur SemEval2013 est de 51,28 % avec un modèle entraîné sur SemCor et WNGT [5].

Figure 3 : Schéma simplifié de Voice2Picto.

**Traitements linguistiques** : « nettoyage du texte » pour pouvoir identifier les pictogrammes associés à chaque terme, avec :

- Une phase de tokenisation, lemmatisation, et d'étiquetage morphosyntaxique,
- La gestion des entités nommées et expressions multi-mots,
- La récupération de chaque image à partir de l'identifiant grâce à une table d'alignement lemmes, sens WordNet et identifiants Arasaac.

## Conclusion

- Voice2Picto **facilite l'accès à la communication** à l'aide de pictogrammes pour un public qui ne connaît pas ou n'utilise pas ce type de support.
- Plateforme simple d'utilisation, rapide, ergonomique, et gratuite.
- Code disponible sur GitHub : <https://github.com/macairececile/Voice2Picto>

## La suite ?

- Améliorer la traduction en définissant un vocabulaire précis et en se basant sur des règles de traduction définies dans une grammaire.
- Gérer les termes non-disponibles sous forme pictographique.
- Évaluer la traduction auprès d'experts.

## Remerciements

Ce travail a bénéficié d'un financement du Fond National Suisse (No. 197864) et de l'Agence Nationale de la Recherche, via le projet PROPICTO (ANR-20-CE93-0005). Ce travail a également bénéficié d'un accès aux moyens de calcul de l'IDRIS au travers de l'allocation de ressources 2022-AD011013625 attribuée par GENCI.

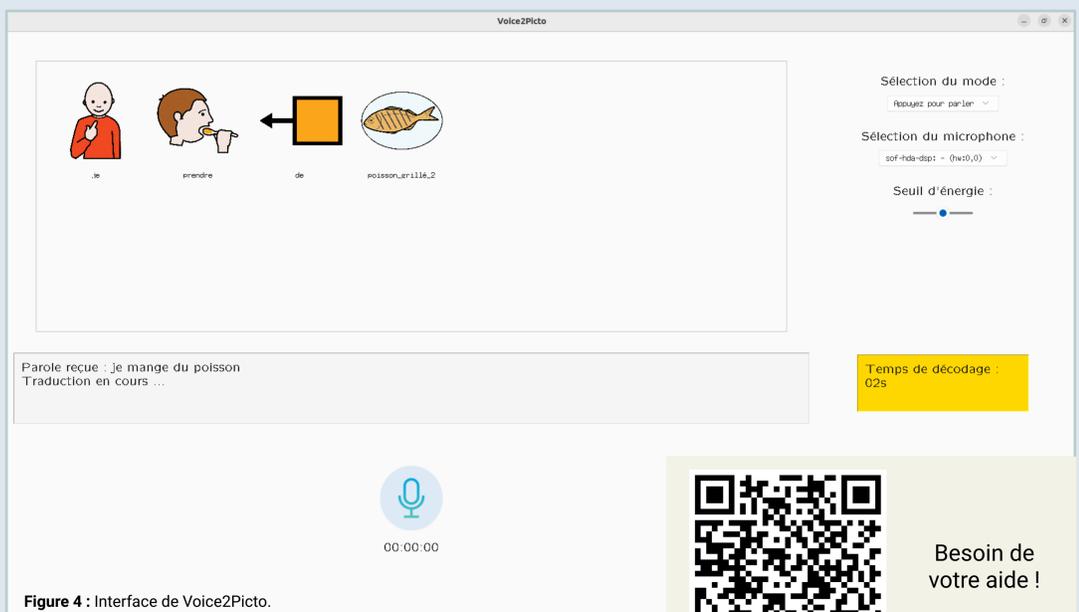


Figure 4 : Interface de Voice2Picto.

## Références

- [1] ARDILA R., BRANSON M., DAVIS K., KOHLER M., MEYER J., HENRETTY M., MORAIS R., SAUNDERS L., TYERS F. & WEBER G. (2020). Common voice : A massively-multilingual speech corpus. In Proceedings of the Twelfth Language Resources and Evaluation Conference, p.4218–4222, Marseille, France : European Language Resources Association.
- [2] CATAIX-NÈGRE E. (2017). Communiquer autrement : Accompagner les personnes avec des troubles de la parole ou du langage. APPRENDRE ET RÉAPPRENDRE. De Boeck Supérieur.
- [3] CROIX-ROUGE (2021). Communiquons autrement - déploiement de la communication alternative améliorée dans les établissements handicap de la croix-rouge française.
- [4] DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers), p. 4171–4186, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : 10.18653/v1/N19-1423.
- [5] LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2020). FlauBERT : Unsupervised language model pre-training for French. In Proceedings of the Twelfth Language Resources and Evaluation Conference, p.2479–2490, Marseille, France : European Language Resources Association.
- [6] MILLER G. A. (1995). Wordnet : a lexical database for english. Communications of the ACM, 38(11), 39–41.
- [7] POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P. et al. (2011). The Kaldi Speech Recognition Toolkit. In IEEE 2011 workshop on automatic speech recognition and understanding, volume CONF : IEEE Signal Processing Society.
- [8] VIAL L., LECOUEUX B. & SCHWAB D. (2019). Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation. In Proceedings of the 10th Global Wordnet Conference, Wroclaw, Poland.